University of Toronto

INF2178

Experimental Design for Data Science

Technical Assignment 4

Instructor

Professor Shion Guha

Submitted by

Lam Hong Kevin Ching

1009243043

2024/04/05

Introduction

The incidence of dementia, characterized by a decline in memory, problem-solving, language, and other cognitive skills that affects a person's ability to perform everyday activities, has been rising globally. With an aging population, understanding the mechanisms and predictors of dementia has become a critical public health priority. Magnetic Resonance Imaging (MRI) plays a pivotal role in this context, offering non-invasive insights into brain structures and functions potentially affected by dementia.

The report will address the following question:

- **Research Question:** How do MRI metrics (nWBV) differ between demented and nondemented groups, controlling for age and education?

MRI metrics, such as Estimated Total Intracranial Volume (eTIV), Normalized Whole Brain Volume (nWBV), and Atlas Scaling Factor (ASF), provide quantitative data on brain anatomy that can be linked to the presence and progression of dementia. However, the relationship between MRI metrics and dementia status is influenced by a myriad of factors, including age and education. Age is a well-known risk factor for dementia, with the prevalence increasing significantly as the population ages. Education, on the other hand, has been suggested to have a protective effect against cognitive decline, possibly due to cognitive reserve theory, which posits that intellectual enrichment throughout life can buffer against brain pathology.

Data Cleaning and Data Wrangling

The studied dataset consists of 294 entries in 16 columns. Initial inspection showed that only two columns (SES and MMSE) contained a small number of missing values, suggesting that the data cleaning effort was minimal within the expected scope of the analysis. I filled in the missing values with the median of these two columns, respectively, and removed the unnamed columns used to index the dataset. The following points summarize the interest observations:

- Group: Demented & Nondemented
- Age: Age in year
- Educ: Education Level
- SES: Socioeconomic status
- MMSE: Mini Mental State Examination
- CDR: Clinical Dementia Rating
- eTIV: Estimated Total Intracranial Volume
- nWBV: Normalize Whole Brain Volume
- ASF: Atlas Scaling Factor

Exploratory Date Analysis

In order to understand the data better and guide the further analysis, I conducted exploratory data analysis before starting to study the research question. First of all, the age distribution (Figure 1) shows a unimodal pattern, with a peak suggesting a concentration of participants in the mid-to-late 70s, followed by a gradual decline as age increases. This indicates that the dataset predominantly consists of older adults, which aligns with the demographic commonly affected by dementia-related conditions. The skewness towards the younger end could imply that younger seniors are underrepresented, while those closer to the peak are more prevalent in the study.

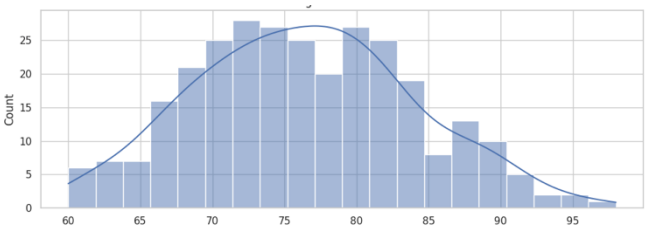

Figure 2: Age Distribution

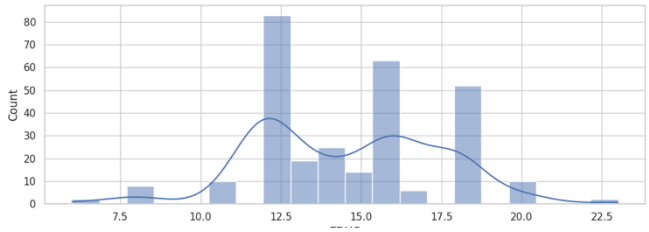

Figure 1: Education Level Distribution

Conversely, the education level distribution (Figure 2) appears to be bimodal, with the majority of participants having around 12 years of education, corresponding to a high school level, and a smaller secondary peak at around 16 years, indicative of a college degree. The bimodal nature reveals two distinct subpopulations within the dataset, suggesting varying levels of educational attainment that may reflect differing socioeconomic backgrounds and potentially different levels of cognitive reserve, which is hypothesized to impact the onset and progression of dementia symptoms.

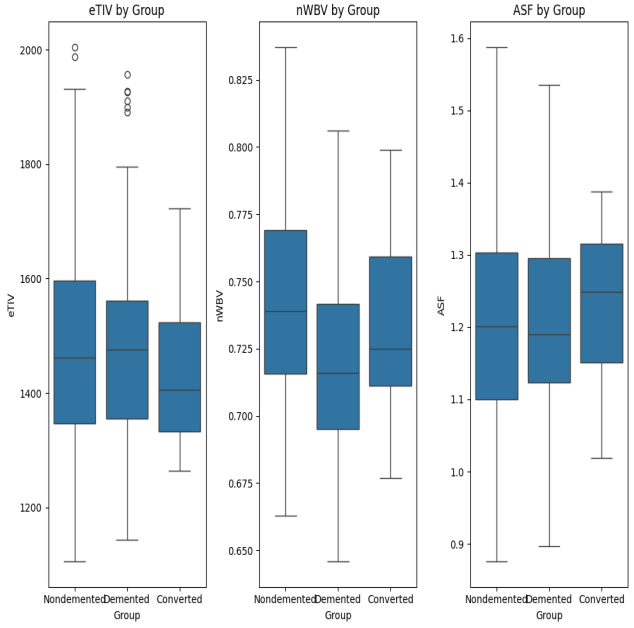

Figure 3: eTIV, nWBV, and ASF Boxplot by Group

Meanwhile, I plot a diagram, which contain three box plots to illustrate the distributions of three MRI metrics—Estimated Total Intracranial Volume (eTIV), Normalized Whole Brain Volume (nWBV), and Atlas Scaling Factor (ASF)—across different patient groups: Nondemented, Demented, and Converted. For eTIV, the central tendency is similar across all groups, with the median eTIV not varying dramatically, suggesting that intracranial volume does not show substantial changes with dementia status. In the case of nWBV, the Nondemented group exhibits higher median values compared to the Demented and Converted groups, which could imply a possible association between brain volume and cognitive health, where higher brain volume might correlate with a nondemented status. The extent of the interquartile range, especially in the Demented group, suggests substantial variation within the group that could be linked to different stages or types of dementia. ASF distributions are relatively consistent across groups, with no group showing a notably different median value.

Mixed-effect ANOVA

Research Question: *How do MRI metrics (nWBV) differ between demented and nondemented groups, controlling for age and education?*

The research question aims to investigate the differences in Normalized Whole Brain Volume (nWBV) between individuals diagnosed with dementia and those without, while accounting for potential confounding variables, specifically age and educational attainment. nWBV is a critical

measure as it reflects brain atrophy, which is often observed in dementia due to the loss of neurons and the connections between them. By controlling for age, the research acknowledges that brain volume naturally decreases with age, and it seeks to discern whether observed differences in nWBV are attributable to dementia rather than normal aging. Education is included as a covariate due to the "cognitive reserve" hypothesis, which suggests that individuals with higher education may be more resilient to the symptoms of dementia, potentially masking the effects on nWBV. This research is fundamental in establishing a clearer link between brain volume and dementia while considering the influence of demographic factors.

	Coef.	*Std.Err*	*z*	*P>|z|*
Intercept	1.033	0.027	37.612	0.000
C(Group)[T.Demented]	-0.022	0.009	-2.555	0.011
C(Group)[T.Nondemented]	0.004	0.009	0.440	0.660
Age	-0.003	0.000	-12.076	0.000
EDUC	-0.001	0.001	-1.133	0.257
Group Var	0.001	0.017		

Table 1: Mixed-effects ANOVA results, with a simplified random effect structure

The mixed-effects ANOVA results (Table 1), using a simplified random effects structure, provide insights into how Normalized Whole Brain Volume (nWBV) varies across different groups while controlling for age and education. The model includes 'Subject ID' as a random effect to account for individual differences. The negative coefficient for the 'Demented' group suggests that, on average, being in the demented category is associated with a decrease in nWBV of 0.022 units compared to the baseline 'Converted' group, and this result is statistically significant ($p = 0.011$). In contrast, the 'Nondemented' group does not show a statistically significant difference in nWBV compared to the 'Converted' group ($p = 0.660$).

Age is negatively associated with nWBV, indicating that as age increases, nWBV tends to decrease; this aligns with typical brain volume reduction observed in aging. The p-value for age in the mixed-effects ANOVA model is indeed less than 0.001, indicating that the effect of age on nWBV is statistically significant. However, education level (EDUC) does not appear to have a significant impact on nWBV in this sample ($p = 0.257$), which might suggest that the variability of nWBV due to education is less pronounced than that due to age or dementia status within this particular dataset. Meanwhile, the group variance is very small (0.001), reflecting the homogeneity of nWBV within subjects, suggesting that the random effect of 'Subject ID' is capturing the within-subject variability effectively.

Overall, the results highlight the importance of dementia status in explaining differences in nWBV, after controlling for age and education, with demented individuals showing a significant decrease in brain volume relative to those who converted. This finding supports the use of nWBV as a potential biomarker in distinguishing between demented and nondemented individuals.

The assumptions underlying the mixed-effects ANOVA model for the nWBV data are critically examined using both a Q-Q plot and Levene’s test. The Q-Q plot (Figure 4), which compares the standardized residuals from the model to a normal distribution, shows that most points follow the

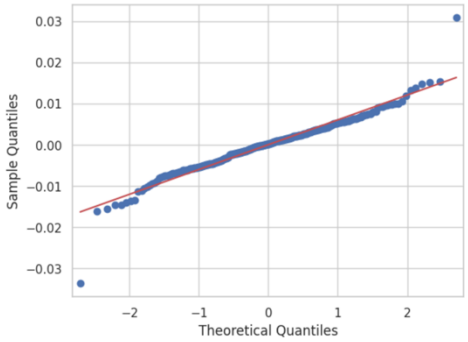

Figure 4: Q-Q Plot of Residuals for nWBV Model

line closely, indicating that the residuals are approximately normally distributed. Levene's test (Table 2) yields a test statistic of approximately 5.4556 with a p-value of 0.0047, which under the conventional alpha level of 0.05, suggesting that the assumption of equal variances is violated. Combining these two assessments, it appears that the data does not significantly violate the assumptions required for conducting a mixed-effects ANOVA.

	statistic	*p-value*
Levene's test	5.4556	0.0047

Table 2: Leven's test result

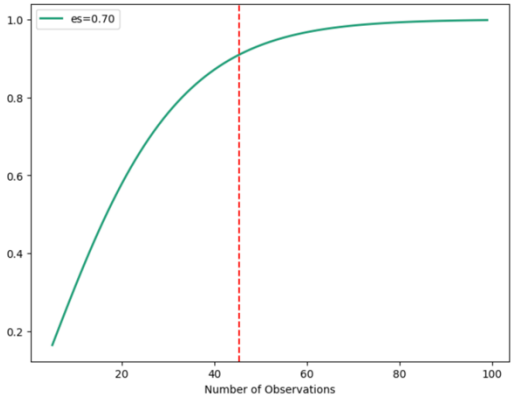

Figure 5: Power of Test

To visualizes the relationship between the number of observations in a study and the power of a t-test, I conduct a power analysis plot (Figure 9). The curve demonstrates that as the sample size increases, the power of the test—our ability to detect an effect if one truly exists—also increases. According to the plot, to achieve a power of 0.91 with an alpha of 0.05 and an effect size of 0.7, the sample size needed is indicated by the vertical dashed line. This number derived from power analysis calculations equal to 45.4507, ensuring precision in determining the required sample size for the given parameters. This means that each group in a two-sample t-test would require roughly 46 participants to have a 91% chance of detecting a true effect of this size.

Conclusion

In conclusion, this comprehensive analysis has elucidated the impact of dementia status on Normalized Whole Brain Volume (nWBV), while carefully adjusting for the confounding effects of age and education. The mixed-effects ANOVA findings reveal a significant decrease in nWBV for the demented group when compared to a baseline, underscoring the potential of nWBV as a marker for brain atrophy associated with dementia. This relationship holds even after accounting for the variance within subjects, as evidenced by the small group variance. Moreover, age is established as a significant factor, with nWBV decreasing as age increases, independent of dementia status. Additionally, the assumption checks show that residuals approximately follow a normal distribution, and that variance is not homogeneously distributed across groups, which might need further investigation using other methodologies. Lastly, the power analysis underscores the adequacy of sample size in detecting the anticipated effect size. This not only bolsters the robustness of the findings but also informs future studies on the requisite sample size for similar research endeavors.